International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-7, Issue-2, February 2017

Feasibility Study of Passive Acoustic and Soft Sensor Based Monitoring of Biological Wastewater Treatment Processes

Sara Nilsson, Mila Harding, Christian Baresel, Anders Björk

Abstract— **Many process parameters in a wastewater treatment plant are expensive, difficult or even impossible to measure online, limiting the possibilities for efficient process monitoring and control. In this work, soft sensors were developed to provide on-line values for a number of parameters, primarily different fractions of phosphate (PO4 and total phosphorous), nitrogen (NO3, NH4 and total nitrogen), organic matter (COD) and suspended solids (TSS), at five different steps of the wastewater treatment process at the R&D-facility Hammarby Sjöstadsverk. The soft sensors were PLS (Partial Least Squares) models predicting the value of the hard-to-measure parameters based on easy-to-measure process parameters that were normally measured on-line or on acoustic data generated by acoustic sensors placed on the tanks of three of the five selected process steps. During a 13-day sampling campaign, data for the soft sensor development and validation were collected by laboratory analysis of the hard-to-measure parameters and combining them with corresponding 5 minute average values of the on-line parameters and the acoustic data. A majority of the soft sensors that were based on acoustic data had comparable or better performance than corresponding models using process data, indicating that data from acoustic sensors are of interest as input variables for soft sensors at WWTPs. The performance of the soft sensors varied significantly and some of them showed promising results. When removing the effect of the laboratory measurement error and the sampling error, 6 out of 26 soft sensor models had a so-called relative true prediction error less than 10% (NO_3 in untreated water, COD, TSS and NO_3 in the first bioreactor, NH_4 in the last bioreactor and TSS in the membrane bioreactor). In combination with the proposed actions for further improvement of the models, the results suggest that soft sensors, that in many cases preferably could be based on acoustic data, is a possible approach to provide WWTPs with on-line process data.**

Index Terms— **Acoustic sensor, Soft sensor, Wastewater treatment**

ABBREVIATIONS

BR	Bioreactor
DO	Dissolved oxygen (mg/L)
COD	Chemical oxygen demand - indirect measure of amount of organic matter (mg/L)
CODf	Dissolved COD (mg/L)
MBR	Membrane bioreactor
NH_4	Ammonium nitrogen - nitrogen in the form of ammonium (mg/L)
NO_3	Nitrate nitrogen - nitrogen in the form of nitrate (mg/L)
N_{tot}	Total nitrogen (mg/L)
PO_4	Phosphate phosphorous - phosphorous in the form of phosphate (mg/L)
P_{tot}	Total phosphorous (mg/L)
TSS	Total suspended solids - solid particles in suspension (mg/L)
TTF	Time to filter - sludge filterability (s)
TTF_{norm}	Time to filter normalized with TSS ($s10^{-4}$/(mg/L))
UF	Ultrafiltration
WWTP	Wastewater treatment plant

Sara Nilsson, IVL Swedish Environmental Research Institute, Stockholm, Sweden, +46 10 7886583
Mila Harding, IVL Swedish Environmental Research Institute, Stockholm, Sweden, +46 10 7886637
Christian Baresel, IVL Swedish Environmental Research Institute, Stockholm, Sweden, +46 10 7886606
Anders Björk, IVL Swedish Environmental Research Institute, Stockholm, Sweden, +46 10 7886572

I. INTRODUCTION

composition and flow rate of wastewater entering a wastewater treatment plant (WWTP) varies greatly during a day as well as between seasons and different weather conditions. Due to the heterogeneity of wastewater and the harsh environment it provides for sensors, many of the parameters relevant for monitoring and control of the process are expensive, difficult or even impossible to measure online, or require substantial maintenance. Some of these parameters are then instead manually analyzed in daily or weekly composite samples providing results sometimes several days or weeks after the samples were taken. Due to the rapidly changing characteristics of the wastewater in combination with the infrequent sampling and the long response time, it is problematic to use the values of manually analyzed parameters to control the plant efficiently. If WWTPs instead had access to real-time values of the important parameters, it could result in a decrease in costs and environmental impact due to more efficient use of chemicals and energy, and also in a decrease in the amount of pollutants that is released to the recipient with the effluent.

One way of providing WWTPs with real-time values for parameters of interest is to use soft sensors, which is the approach used in this article. A soft sensor is a virtual sensor that predicts the value of a parameter whose value is unknown, e.g. a parameter that is hard to measure online, solely based on values of other parameters whose values are known, e.g. parameters that are easier to measure online.

The multivariate statistical regression method that was found suitable for developing the soft sensor models in this work is PLS (Partial Least Squares or Projection to Latent Structures) [1],[2]. With PLS, the aim is to establish the relationship between input (x) variables, and output (y) variables. A PLS model is calculated in such a way that it

describes as large portion as possible of variance in the data, at the same time as it maximizes the covariance between the x-variables and the y-variables. The final result is an equation expressing y as a linear combination of the x-variables. There are several relevant papers regarding process data based soft sensors in WWTPs. In an early work, Mujunen et al. used PLS to model a wastewater plant with activated sludge treatment[3]. Rosén et al. presented solutions and challenges regarding use of multivariate models in waste water treatment [4]. A recent work with similar scope is the work by Haimi[5].

Another option is to use acoustical spectra as input variables for the soft sensors rather than ordinary process data. For a general introduction to the use of vibration measurements to predict the properties of different fluids, as well as a background on vibrations, measurement technology, signal processing, multivariate analysis and applications, see [6]. There are several industrial applications for acoustic spectra as a basis for determining parameters for process monitoring or control. Some examples are to use it for monitoring of oil production wells[7], for prediction of content of different types particulate material (e.g. alumina, PVC, sand) in a pneumatic transport-tube [8, predicting content of oil, glycol and paper-pulp constituents in water by measurements on a constriction in pipe-line {Esbensen, 1999 #336], in industrial plastic granulation process utilizing microphones [9], in mechanical paper pulp production for measurements of pulp quality[10],[11],[12] and for determining food textural properties of snacks[13]. As shown by these references, vibration measurements can be used to determine properties of different fluids, and could presumably serve as input-variables for soft sensors in form of acoustic spectra generated by acoustic sensors installed in a wastewater treatment process.

This article covers the development and evaluation of soft sensors for a number of parameters in different process steps in a wastewater treatment process. The softs sensors were based on ordinary online process variables as well as acoustic spectra from accelerometers mounted on process reactors.

II. MATERIALS AND METHOD

The soft sensors were developed for the pilot scale WWTP Hammarby Sjöstadsverk in Nacka, Sweden[14], for which a process overview is presented below.

Data for the soft sensors were generated by laboratory analysis of samples collected during a sampling campaign and by gathering corresponding process data and acoustic data from the control system. This is described in more detail in the sections for "Sampling and data collection" and "Laboratory analysis". Soft sensor PLS models were then calculated and thereafter externally validated, procedures that are covered by the "Modelling and validation" section.

A. *Process overview*

The study was performed at line 1 at the pilot WWTP Hammarby Sjöstadsverk (Fig. 1). Line 1 is a pilot scale membrane bioreactor (MBR) of thefuture Henriksdal plant, Stockholm's largest WWTP, and has a capacity corresponding to 250 connected persons. The first process step is pre-sedimentation, where phosphorous containing sludge settles to the bottom of the basin after the addition of a coagulant. In the following biological treatment, consisting of three unaerated bioreactors (BR1-3) and three aerated bioreactors (BR4-6), nitrogen is removed from the water in the form of nitrogen gas by the microorganisms in the continuously recirculating sludge. Water is re-circulated from the aerated to the unaerated bioreactors to facilitate the nitrogen removal. In the membrane bioreactor (MBR), the sludge is separated from the water by a submerged ultrafilter (UF). The majority of the sludge is re-circulated to retain a high concentration of microorganisms in the biological treatment step, and the water passing through the filter is ready to be released to the recipient.

B. *Sampling and data collection*

During 13 days in October 2014, grab samples were collected from incoming (untreated) wastewater, BR 1, BR5, BR6 and from the MBR(the unit with an UF). Which parameters to analyze were selected based on their relevance in each process step. Samples for analysis of PO_4, NH_4, NO_3 and COD_f were manually collected every fourth hour between 08:00 and 16:00 on weekdays and were filtered through a 0.45 µm syringe filter within 2 minutes after collection. Samples for analysis of P_{tot}, N_{tot}, TSS, COD and sludge filterability (TTF) were collected with automatic samplers (6712 Portable Sampler, Isco) every fourth hour around the clock every second day. The samples were stored in the partially ice-filled insulated samplers and/or in a +4^{0}C fridge until analyzed.

In addition to the standard online sensors, the process line was also equipped with acoustic sensors (Ceramic Shear Integrated Electronic Piezoelectric Accelerometer Type 8714B100M5, Kistler) on BR1, BR5 and MBR.

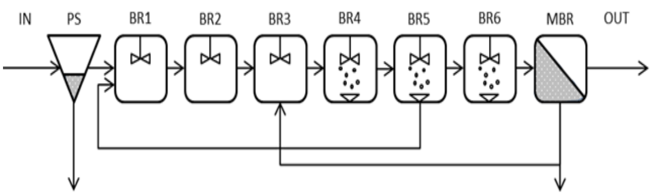

Figure 1. Schematic overview of the treatment line at HammarbySjöstadsverk used for development of the soft sensors. The incoming water is named IN and effluent is named OUT. The line consists of a pre-sedimentation (PS), three unaerated bioreactors (BR1-3), three aerated BR (BR4-6) and a membrane bioreactor (MBR) with the submerged UF

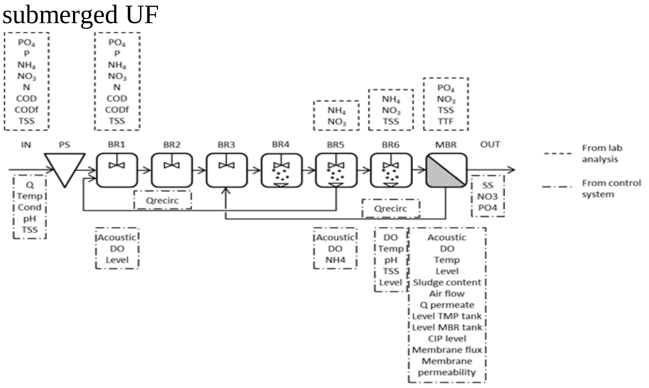

Figure 2. . Data collected from the different process steps during the sampling campaign: parameters measured with lab analysis (dash outline) and parameters from the control system (dash dot outline).

Two cDAQ 9181 chassis with respectively one NI 9234 IEPE and AC/DC Analog Input module (from National instruments) were used to collect data from the accelerometers. LabVIEW was used as programing language

to gather data from the modules, compute power spectra from each accelerometer and save data to the PostgreSQL database. The frequency range was 0-25.6 kHz for the spectra, with 1024 frequency bins.

Values from the sensors were stored in the database every minute together with other process parameters. 5 minute average values were calculated for each of the parameters, and the average values corresponding to the manual sampling times were used for the modelling. The parameters available from each process step are summarized in Fig. 2.

A. *Laboratory analysis*

The concentrations of PO_4, P_{tot}, NH_4, NO_3, N_{tot}, COD and COD_f were determined in mg/L with cuvette tests (WTW) that were analyzed with a spectrophotometer (photoLAB 6600 UV-VIS, WTW). TSS was measured in mg/L by filtering the sample through a standard 55 mm GF/F glass fiber filter with 1.6 micrometer pores that had been previously dried in 105˚C. After the filtration, the filter was dried in 105˚C over night. TSS was calculated by dividing the difference in filter weight before and after filtration with the volume of sample that was filtered. The sludge filterability was measured in terms of time to filter (TTF), i.e. the time in seconds required for a certain volume of sample to pass a 90 mm glass microfiber filter with 1.5 micrometer pores (Grade 934-AH RTU, Whatman, GE Healthcare) with a vacuum of 15 mmHg. This was done according to the method specified in GE Water & Process Technologies, 2009. Also, a TTF value normalized with TSS was calculated according to (1).

$$TTF_{norm} = \frac{TTF}{TSS/10^4} \qquad (1)$$

B. *Model calibration and validation*

Soft sensor models were developed for all manually analyzed parameters. Before calculating the models, data were split into a calibration set for the calculation of each model, and a validation set for external validation of the model. The first 1/6 and last 1/6 of the data were selected as validation set, and the rest was used as calibration set. All data, except for the acoustic data, were centered and scaled to unit variance before modelling.

PLS models were calculated for the data in the calibration set. To improve the models, x-variables that did not contribute to the models were excluded. The decision of which x-variables to exclude was based on each variable's VIP-value, which reflects the extent to which the variable explains X and correlates to Y. [15],

To evaluate the models and select the best model for each parameter, a combination of cross validation and permutation testing was used. The cross-validation gave an estimate of the predictive power of the models and was made with 7 cross validation groups, where the first 1/7 of the observations formed the first group, the second 1/7 of the observations formed the second group and so on. The response permutation testing was then used to decrease the risk of selecting models that were overfitted to the calibration data. For more information, see [15]. The best model for each parameter in each sampling point was then externally validated with the data in corresponding validation set. The statistical measures used for evaluation of the models are presented in (2) – (8) in Appendix.

The software used for the modelling was SIMCA v14.1(MKS Data Analytics Solutions).

Table 1. Number of observations available for the parameters measured in each sampling point.

	IN	**BR1**	**BR5**	**BR6**	**MBR**
NO_3	28	28	28	28	28
NH_4	28	28	28	28	
PO_4	28	28			28
CODf	28	27			
TSS	42	42		42	36
P_{tot}	42	42			
N_{tot}	42	42			
COD	42	42			
TTF					36
Acoustic spectra	57	57	57	57	57
Process parameters	57	57	57	57	57

III. RESULTS AND DISCUSSION

A. *Sampling and data collection*

The sampling campaign resulted in 27 to 42 values for each analyzed parameter in each sampling point (**Error! Reference source not found.**). 5 minute average values were calculated for corresponding process parameters and acoustic signals from the control system. Thus, the number of observations was limited for both the modelling and the external validation.

B. *Model calibration*

Several models were then developed for each parameter. The best model for each parameter with respect to Q^2 and $RMSE_{cv}$, and that still passed the permutation testing was selected for further validation. The selected models consisted of a wide range of different x-variables and were of very varying qualitywith respect to R^2, Q^2 and $RMSE_{cv}$. In the cases where both process parameters and acoustic data were available, models containing only one of the two data types were prioritized if the performance of the models were comparable. The properties of the selected models are summarized in **Error! Reference source not found.**, and the specific x-variables that were used in each model are presented in Table 4 in Appendix.

C. *Model validation*

The best model for each parameter was externally validated with data from the beginning and the end of the sampling campaign. The external validation was evaluated based on the prediction error (RMSEP and rel RMSEP) and the true relative prediction error ($RMSEP_{true}$ and $relRMSEP_{true}$).

To calculate $RMSEP_{true}$ and $relRMSEP_{true}$ according to (7) and (8) in Appendix, estimations of measurement error and sampling error were needed. For the parameters analyzed with cuvette tests, the measurement error was defined as the measurement uncertainty specified for each test by the manufacturer, and for TSS and TTF it was defined as 5% of the average value for each parameter in the training set. The sampling error was assumed to be 5% of the average value for each parameter in the training set.

Pos	Y	X	Samples	A	R^2	Q^2	Y range	$RMSE_{cv}$
IN	P_{tot}	4 process variables	28	1	0.611	0.529	1.8 -10.3	1.28
	N_{tot}	4 process variables	28	1	0.417	0.355	4.3 - 68	11.8
	COD	4 process variables	28	1	0.557	0.369	245 - 857	132
	TSS	3 process variables	28	1	0.263	0.113	98 - 368	77
	PO_4	5 process variables	19	2	0.915	0.837	1.0 - 4.6	0.45
	NO_3	4 process variables	19	2	0.751	0.491	0.06 - 0.88	0.17
	NH_4	5 process variables	19	1	0.855	0.812	4.3 - 43.2	4
	CODf	4 process variables	19	1	0.66	0.612	63 - 347	49
BR1	P_{tot}	513 acoustic variables	28	3	0.987	0.851	133 - 193	9.9
	N_{tot}	513 acoustic variables	28	1	0.377	-0.024	120 - 240	27.1
	logCOD	513 acoustic variables	28	4	0.995	0.823	5620 - 10510	690
	TSS	513 acoustic variables	28	3	0.99	0.897	5080 - 7262	289
	PO_4	513 acoustic variables	19	4	0.995	0.747	0.09 - 0.49	0.08
	NO_3	513 acoustic variables	19	1	0.372	-0.1	0.01 - 0.94	0.3
	NH_4	513 acoustic variables	19	3	0.944	0.842	0.6 - 9.6	0.23
	CODf	513 acoustic variables	18	2	0.909	0.766	47 - 102	8.8
BR5	NO_3	184 acoustic variables	19	3	0.906	0.754	0.16 – 6.1	1
	NH_4	513 acoustic variables	19	4	0.996	0.815	0.02 – 1.7	0.37
BR6	TSS	4 process variables	27	1	0.7	0.647	6419 - 8381	367
	NO_3	3 process variables	19	1	0.656	0.573	0.05 – 4.9	1.11
	NH_4	2 process variables	19	1	0.323	0.264	0.018 -1.597	0.559
MBR	TSS	14 process variables	24	3	0.958	0.89	7742 – 10033	336
	PO_4	10 process variables	19	1	0.561	0.462	0.07 – 0.45	0.08
	NO_3	12 process variables	19	2	0.761	0.643	0.07 – 6.5	0.98
	TTF	24 process variables	24	3	0.867	0.64	40 - 54	2.37
	TTF_{norm}	182 acoustic variables	24	5	0.912	0.584	46 - 58	1.88

Table 2. Properties of the best PLS-model for each parameter. Pos – position of the soft sensor, Y – parameter, X – number and type of x-variables. Samples - number of samples that the model is based on, A – number of principal components in the PLS model, , R^2 – variance explained by the model, Q^2 – estimate of predictive ability, Y range – range of the parameter in the calibration set (in mg/L, s or sL/mg), RMSE.

Pos	Y	Y range	ME*	SE*	RMSEP	**relRMSEP**	**$RMSEP_{true}$**	**$relRMSEP_{true}$**
IN	P_{tot}	3.6-8.9	0.4	0.3	1.51	17.8	1.43	16.8
	N_{tot}	22 - 73	5	2.29	11.5	18.1	10.1	15.9
	COD	413 - 997	29	29.5	142	23.2	136	22.2
	TSS	150 - 394	11.4	11.4	75	27.8	73.2	27.1
	PO_4	2.3 – 5.1	0.4	0.16	0.78	21.7	0.65	18
	NO_3	0.08 – 0.61	0.3	0.02	0.17	20.7	0	0
	NH_4	16.8 – 40.8	1.9	1.26	5.8	14.9	5.33	13.7
	CODf	116 - 339	29	10.8	43	15.1	29.9	10.5
BR1	P_{tot}	137 - 191	0.06	8.03	11.8	19.7	8.65	14.4
	N_{tot}	170 - 240	0.5	9.34	31.6	26.3	30.2	25.2
	logCOD	6250 - 8350	29	348	581	11.9	465	9.5
	TSS	5587 - 7089	298	298	406	18.6	0	0
	PO_4	0.22 - 0.72	0.06	0.01	0.27	67.5	0.26	65.8
	NO_3	0.05 - 0.40	0.3	0.01	0.25	26.9	0	0
	NH_4	2.3 - 8.4	0.2	0.26	1.21	13.4	1.16	12.9
	CODf	64 - 116	7	3.89	15.2	27.6	12.9	23.5
BR5	NO_3	0.32 – 6.3	0.3	0.16	2.1	35.4	2.07	34.9
	NH_4	0.014 – 1.076	0.05	0.02	0.53	31.5	0.53	31.4
BR6	TSS	7113 - 8027	361	361	494	25.2	0	0
	NO_3	0.06 – 5.1	0.3	0.12	1.86	38.4	1.83	37.8
	NH_4	0.016 – 1.005	0.05	0.02	0.137	8.7	0.12	7.9
MBR	TSS	8733 - 11503	437	437	310	13.5	0	0
	PO_4	0.1 – 0.38	0.06	0.01	0.08	21.1	0.05	13.8
	NO_3	0.11 – 5.0	0.3	0.14	1.39	21.6	1.35	21
	TTF	45 - 71	2.24	2.24	17	121.4	16.7	119.3
	TTF_{norm}	50 - 67	2.56	2.56	16.9	140.8	16.5	137.6

Table 3. Results from the external validation. Pos – position of the soft sensor, Y – parameter, Y range – range of the parameter in the prediction set (in mg/L, s or sL/mg), ME – measurement error of the laboratory analysis, SE – sampling error, RMSEP – prediction error of the model for external validation, relRMSEP - relative RMSEP, $RMSEP_{true}$ – RMSEP adjusted for measurement error and sampling error, $relRMSEP_{true}$ - relative RMSEP adjusted for measurement error and sampling error.

Out of 26 soft sensors, 4 models had a relRMSEP of less than 15% (NH_4 in untreated water based on process data, COD in the first bioreactor based on acoustic data, NH_4 in the last bioreactor based on process data and TSS in the membrane bioreactor based on process data). 12 models had a $relRMSEP_{true}$ of less than 15%, out of which 6 models had a $relRMSEP_{true}$ of less than 10% (NO_3 in untreated water based on process data, COD, TSS and NO_3 in the first bioreactor based on acoustic data, NH_4 in the last bioreactor based on process data and TSS in the membrane bioreactor based on process data). This indicates that the measurement error and the sampling error in many cases significantly affected the prediction error of the models. In some cases, the sum of ME^2 and SE^2 even exceeded $RMSEP^2$ (indicated by $RMSEP_{true}$ value of 0 in **Error! Reference source not found.**, where the results are presented in more detail).

A. Concluding discussion

For a majority of the parameters at the positions where acoustic sensors were installed (BR1, BR5, MBR), the models using acoustic data had comparable or better performance than corresponding models using process data. Thus, installing acoustic sensors in the process steps where

acoustic data were not available could improve the soft sensors. This also indicates that acoustic measurements could have the potential to be used as input to soft sensors for WWTPs in general.

Due to the relatively few observations available, the conclusions that can be drawn from this study are limited. Since the composition of the incoming wastewater varies greatly between seasons and different weather conditions, more sampling campaigns should be done. Preferably, they should be spread out over at least one year to generate data that is representable enough to be able draw more extensive conclusions

s about the suitability of soft sensors as a possible method to generate on-line data for wastewater treatment.

One more aspect to take into consideration when interpreting the results from the external validation is that the amount of rainfall varied considerably during the sampling campaign, which significantly affected the composition of the wastewater. Compared to if there had been a more constant amount of precipitation during the sampling campaign, the changing weather conditions resulted in a dataset that represented a relatively wide range of different wastewater compositions, which is positive for the range for which the models are valid. But, it also increased the risk that the range of compositions in the external validation set was not covered by the calibration set, which results in that the external validation indicates that the predictive ability of the models is lower than if the validation set would have been representable for the calibration set.

However, with this in mind, some of the soft sensors showed promising results, especially NO_3 in incoming wastewater, COD, TSS and NO_3 in BR1, TSS and NH_4 in BR6 and TSS in the MBR, all of which had a relative true prediction error less than 10%.

Moreover, the models can probably be further improved by optimizing the calculation of the acoustic spectra and the signal processing, for example by testing different spectral algorithms and weighting windows (e.g. Tukey) before applying fast Fourier transform to produce the spectra. It would also be interesting to evaluate other types of accelerometers, possibly with a narrower bandwidth and better sensitivity.

IV. CONCLUSIONS

Soft sensors were developed for 26 parameters at five process steps at the pilot scale WWTP Hammarby Sjöstadsverk in Sweden. A number of soft sensors showed a relatively good predictive ability, which indicates that soft sensors have the potential to provide WWTPs with on-line values for parameters relevant for process monitoring and control.

For the majority of the parameters, the soft sensors that were based on acoustic data had comparable or better performance than corresponding models based process data. This brings us to the conclusion that data from acoustic sensors are of interest as input variables for soft sensors at WWTPs.

It is also our belief that the soft sensors can be further improved by calibrating them with data generated during a longer period of time. This could reduce the prediction errors and as expand the validity domains of the models, and/or by improving the acoustic data by optimizing the calculation of the acoustic spectra and the signal processing or using other types of accelerometers. This further strengthens the conclusion that soft sensors is a promising approach for WWTPs.

APPENDIX

A. Quality of multivariate statistical models

The quality of the PLS models can be represented by the following measures:

R^2 -the part of the variance explained in the calibration data, i.e. a measure of how well the model fits the calibration data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-y)^2} \qquad (2)$$

where *(y - ŷ)*refers to the fitted residuals for the observations in the calibration set and *n* refers to the number of samples.

Q^2 - an estimate of the predictive ability of the model and is calculated by cross-validation. If Q^2 is 1, the model predicts the data perfectly.

$$Q^2 = 1 - \prod\left(\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}\right)_a \qquad (3)$$

where*(y - ŷ)*refers to the predicted residuals for the observations in the calibration set during cross-validation, *n*

refers to the number of samples and *a* refers to the principal components.

RMSEcv (root mean square error of cross validation) – an estimate of the predictive power of the model based on cross validation. It has the same unit as the y-variable.

$$RMSEcv = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (4)$$

where *(y - ŷ)*refers to the predicted residuals for the observations in the calibration set during cross-validation and *n* refers to the number of samples.

RMSEP (root mean square error of prediction) - a measure of the predictive power of a model. It has the same unit as the y-variable.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (5)$$

where *(y - ŷ)*refers to the predicted residuals for the observations in the external validation data set and *n* refers to the number of samples.

relRMSEP - a measure of the relative predictive power of a model. Given in %.

$$relRMSEP = 100 \frac{RMSEP}{y_{max} - y_{min}} \quad (6)$$

where *(y - ŷ)* refers to the predicted residuals for the observations in the external validation data set, *n*refers to the number of samples and y_{max}-$y_{min}$to the range of the y-variable in the calibration set.

RMSEP$_{true}$ - a measure of the prediction error of the model after adjusting for the measurement error and sampling error. It has the same unit as the y-variable.

$$RMSEP_{true} = \sqrt{RMSEP^2 - ME^2 - SE^2} \quad (7)$$

where *RMSEP* refers to the prediction error of the model, *ME* to the measurement error and *SE*to the sampling error.

relRMSEP$_{true}$- a measure of the relative prediction error of the model after adjusting for the measurement error and sampling error. Given in %

$$rel\,RMSEP_{true} = 100 \frac{\sqrt{RMSEP^2 - ME^2 - SE^2}}{y_{max} - y_{min}} \quad (8)$$

where *RMSEP* refers to the prediction error of the model, *ME*to the measurement error, *SE*to the sampling error, y_{max}-y_{min} to the range of the y-variable in the calibration set.

B. Variables in models

Pos	**Y**	**X**
IN	P_{tot}	4 process variables (Temp, Cond, pH, TSS)
	N_{tot}	Q, Temp, pH, SS
	COD	Temp, Cond, pH, TSS
	TSS	Temp, Cond, pH
	PO_4	Q, Temp, Cond, pH, TSS
	NO_3	Q, Temp, Cond, TSS
	NH_4	Q, Temp, Cond, pH, TSS
	CODf	Q, Temp, Cond, TSS
BR1	P_{tot}	513 acoustic variables
	N_{tot}	513 acoustic variables
	logCOD	513 acoustic variables
	TSS	513 acoustic variables
	PO_4	513 acoustic variables
	NO_3	513 acoustic variables
	NH_4	513 acoustic variables
	CODf	513 acoustic variables
BR5	NO_3	184 acoustic variables
	NH_4	513 acoustic variables
BR6	TSS	logDO, pH, Temp, Qin
	NO_3	pH, Temp, Level
	NH_4	logDO, logQin

MBR	TSS	Level, Sludge content, Air flow, $Q_{premeate}$, Levels TMP tanks, Level MBR tanks, CIP levels TMP tanks, Membrane fluxes, NO_3 out, PO_4 out
	PO_4	Level, Sludge content, Q_{recirc}, $Q_{premeate}$, Level MBR tanks, Membrane flux, Membrane permeability, NO_3out, Qin
	NO_3	Level, DO, Sludge content, Air flow, $Q_{premeate}$, Levels TMP tanks, Level MBR tanks, CIP levels TMP tank, Line stat, Membrane flux, PO_4out
	TTF	Level, DO, Sludge content, Q_{recirc}, Air flows, $Q_{premeate}$, Levels TMP tanks, Level MBR tanks, CIP levels TMP tanks, Line stat, membrane fluxes, Membrane permeabilities, NO_3out, PO_4out, Qin
	TTF_{norm}	182 acoustic variables

ACKNOWLEDGMENT

The authors would like to thank the staff at Hammarby Sjöstadsverk fortheir help at site and Nils Andersson for the installation work at the plant.

The economic support from the Foundation for IVL Swedish Environmental Research Institute (SIVL).

REFERENCES

[1] Geladi, P. and B.R. Kowalski, Partial Least-Squares Regression : A Tutorial. *Anal. Chim. Acta*. 185, 1986, pp. 1-17.

[2] Martens, H. and T. Naes, *Multivariate Calibration*, in *Multivariate Calibration*N.Y, John Wiley & Sons.1989, pp. 1-700.

[3] Mujunen, S.-P., et al., Modeling of activated sludge plants treatment efficiency with PLSR: a process analytical case study. *Chemometr. Intellig. Lab. Syst.*41(1), 1998, pp. 83-94.

[4] Rosen, C., J. Rottorp, and U. Jeppsson, Multivariate on-line monitoring: challenges and solutions for modern wastewater treatment operation. *Wat. Sci. Technol.*47(2), 2003, pp. 171-179.

[5] Haimi, H., Data-derived soft sensors in biological wastewater treatment. . 2016, Aalto University, Finland.

[6] Björk, A., Chemometric and signal processing methods for real time monitoring and modeling using acoustic sensors Applications in the pulp and paper industry. Doctorial Thesis. 2007, Royal Institute of Technology School of Chemical Science and Engineering Department of Chemistry Division of Analytical Chemistry: Stockholm. pp. 113.

[7] Hope, B.R., *A Method for Recording Multi-Phase Flows Through a Transport System*: International Application PCT, WO 89/05974 (29 June 1989), NO Pat. Appl. 875326 Prior. 18 December 1987. 1989.

[8] Esbensen, K.H., et al., Acoustic chemometrics—from noise to information. *Chemometr. Intellig. Lab. Syst.*44, 1998, pp. 61–76.

[9] Halstensen, M., P. de Bakker, and K.H. Esbensen, Acoustic chemometric monitoring of an industrial granulation production process--a PAT feasibility study. *Chemometr. Intellig. Lab. Syst.*84(1-2), 2006, pp. 88-97.

[10]Björk, A., L.-G. Danielsson, and A. Gannå. Does It Sound Like The Right Freeness Today? Or Applied Near Real-Time Acoustic Chemometrics For Measurements of Canadian Standard Freeness. in

International Mechanical Pulping Conference. Helsinki, 2001, The Finnish Paper Engineers' Association.

[11]Björk, A. and L.-G. Danielsson, Spectra of wavelet scale coefficients from process acoustic measurements as input for PLS modelling of pulp quality. *Journal of Chemometrics***.** 16(8-10), 2002, pp. 521-528.

[12]Björk, A. and L.G. Danielsson, Extraction of distribution curvesfrom process acoustic measurements on a TMP-process*. Pulp & Paper Canada***.** 105(11), 2004, pp. T260-T264.

[13]Bruwer, M.J., J.F. MacGregor, and W.M. Bourg, Soft Sensor for Snack Food Textural Properties Using On-Line Vibrational Measurements. *Indust. Eng. Chem. Res.*46(3), 2007, pp. 864-870.

[14]Baresel, C. (2016-10-20). *Hammarby Sjöstadsverk* Available: http://hammarbysjostadsverk.se/Sjostadsverket/english/hammarby-sjostadsverk.html.

[15]L. Eriksson, E.J., N. Kettaneh-Wold, S. Wold, *Multi- and Megavariate Data Analysis. Principles and Applications*: Umetrics Academy. 2001.

Sara Nilsson has a MSc in biotechnology from Umeå University, Umeå, Sweden

Sara has a background in environment, biotechnology and multivariate data analysis and has been working for more than 10 years with research, assignments and project management within the environmental field with focus on data analysis, process modelling, wastewater, drinking water and environmental technology

Mila Harding is a skilled expert in water purification and has a specific focus on analytics, experimental design and pilot operation. She has an Advanced Higher Vocational Education in Environmental engineering from Stockholm Environmental center. She has worked with water- and wastewater technologies since 2003 at Stockholm Water, Veolia Water and has since 2011 been one of the driving professionals at the R&D-facility Hammarby Sjöstadsverk. Mila has been working with water related innovations and technologies in a number of different projects. She has been part of several technical reports about developing and optimizing processes with the aim to reuse treated wastewater and the removal of pharmaceutical residues.

Christian Baresel is the manager and coordinator for R&D-facility Hammarby Sjöstadsverk on innovative wastewater treatment technologies. PhD in aquatic engineering from KTH-Royal Institute of Technology, Stockholm. Author/co-author of more than 11 peer-reviewed scientific publications; 9 conference papers, and numerous of technical reports. He is a senior project manager and expert in sewage treatment technologies, bioenergy production from sewage, separation technologies and removal of micropollutants.

Anders Björk has PhD in process analytical chemistry and MSc in Chemical Engineering and Chemistry, both from KTH-Royal Institute of Technology, Stockholm, Sweden. Author/co-author of more than 8 peer-reviewed scientific publications; 7 conference papers, and a number of technical reports. Articles within process chemometrics, analytical chemistry, environmental impact in supply chains, particle emissions from trains etc. Current position as senior scientist and project leader, IVL Swedish Environmental Research Institute Ltd. He has also commission of trust as Employee company board representative IVL Swedish Environmental Research Institute Ltd. since 2014 and as Member of the research council of IVL Swedish Environmental Research Institute Ltd. since 2012.

He has broad research profile with the aim to realizing sustainability and resource efficiency by models, sensors, automation and IT. Industrial technologies, sustainability and research in different fields. Summarized by the concept Industry 4.0 as well as term process intensification PI. He has worked within these areas energy analysis, process control and automation, software development utilizing service oriented architecture, environmental product declarations, environmental footprint calculations, interaction between environment and health, cleantech and sensor technologies. Multivariate analysis of big data has been a core area since 1997 this also includes work on what is now called deep learning. A part of the work has been since 2007 handling requests from researchers while keeping security and maintaining IVLs IT systems and different automations system at Hammarby Sjöstadsverk as well as performing research